IDENTIFICATION OF A REGION OF THE MAJOR SURFACE GLYCOPROTEIN (*MSG*) GENE OF HUMAN *PNEUMOCYSTIS CARINII*

REFERENCE TO RELATED CASES

This is a divisional of co-pending U.S. Patent Application No. 09/762,724, filed February 9, 2001, which is the United States National Phase of International Application Number PCT/US99/18750, filed August 17, 1999, which claims the benefit of U.S. Provisional Application No. 60/096,805, filed August 17, 1998. Each of the foregoing applications is incorporated herein in its entirety.

FIELD OF THE INVENTION

This invention relates to methods for detecting *Pneumocystis carinii* infection in humans, specifically to such methods that involve polymerase chain reaction or other amplification of nucleic acid sequences that encode a *Pneumocystis carinii* sp. f. hominis protein.

BACKGROUND OF THE INVENTION

Pneumocystis carinii is an important life threatening opportunistic pathogen of immunocompromised patients, especially those with human immunodeficiency virus (HIV) infection. Conventional diagnosis of *Pneumocystis carinii* pneumonia (PCP) involves analysis of a tissue sample or oropharyngeal secretion sample for the presence of a *P. carinii* organism through staining and microscopic examination. Sample acquisition techniques have included such invasive methods as transbronchial biopsy, percutanenous lung biopsy, or open lung biopsy. Each of these techniques is fraught with possible complications and requires significant time and expense. In the mid 1980's, bronchoalveolar lavage (BAL) was introduced as a less invasive, less expensive, and less complication-prone technique for acquiring samples to be used in PCP diagnosis (Ognibene *et al., Am. Rev. Respir. Dis.* 129:929-932, 1984). However BAL, coupled with bronchoscopy, still required special equipment and facilities, as well as the time of a physician and technician. Simpler still, it is now known that the *Pneumocystis* organism

can also be detected in induced sputum samples (Bigby *et al., Am. Rev. Respir. Dis.* 133:515-518, 1986; Kovacs *et al. NEJM* 318:589-593, 1988).

Advances also have occurred in the techniques used to detect the *Pneumocystis* organism in tissue and oropharyngeal secretion samples. Direct microscopic examination of clinical samples stained with, for instance, Giemsa stain or toluidine blue O, requires time-consuming sample preparation and subsequent examination by specially trained and experienced microscopy technicians (see, for instance, Bigby *et al., Am. Rev. Respir. Dis.* 133:515-518, 1986). This procedure has been somewhat simplified and rendered more amenable to mechanization through the use of monoclonal antibodies in detection of *P. carinii* antigens in clinical samples (Kovacs *et al., NEJM* 318:589-593, 1988). A few groups have used oligonucleotide probes complementary to *P. carinii* nucleotide sequences to detect the organism through hybridization, as in U.S. Pat. No. 5,164,490 (the Santi patent).

Polymerase chain reaction (PCR) -mediated amplification of DNA or RNA-encoding sequences has been used to diagnose various diseases including leprosy (Santos *et al., J. Med. Microbiol.* 46:170-172, 1997) and PCP. This technique exhibits increased sensitivity over simple probe hybridization methods. Primers complementary to sequences encoding *P. carinii* mitochondrial or chromosomal ribosomal RNA (rRNA) have been used to amplify *Pneumocystis*-specific DNA sequence, as in Wakefield *et al. Mol. Biochem. Parasit.* 43:69-76, 1990; Wakefield *et al. Lancet* 336:451-453, 1990; Lipschik *et al. Lancet* 340:203-206, 1992; WO 91/19005; and U.S. Pat. Nos. 5,519,127 (the Shah patent), 5,593,836 (the Niemiec patent) and 5,776,680 (the Leibowitz patent).

Other recent research advances relate to elucidating the molecular mechanisms involved in *P. carinii* infection. A great deal of interest has focused on the major surface glycoprotein (MSG; also called glycoprotein A) of *P. carinii*, because it is considered to be both a virulence factor and a target of host immune responses. MSG is the most abundant protein expressed on the surface of *P. carinii*, as assessed by Coomassie blue staining. It appears to play a critical role in the pathogenesis of pneumocystosis, possibly by acting as an attachment ligand to lung cells. MSG is also a target of both humoral and cellular immune responses by the host.

Multiple genes encode the *MSG* of rat-*P. carinii*, and different *MSGs* may be expressed in the lung of a rat infected with *P. carinii* (Angus *et al., J. Exp. Med.* 183:1229-1234, 1996; Kovacs *et al., J. Biol. Chem.* 268:6034-6040, 1993). Similarly, multiple genes encode the *MSG* of *P. carinii* infecting ferrets and mice (Haidaris *et al., DNA Res.* 5:77-85, 1998; Haidaris *et al., J. Infect. Dis.* 166:1113-1123, 1992). Additional studies have shown that there is a single genomic site for expression of rat *MSG* variants (Edman *et al., DNA Cell Biol.* 15:989-999, 1996; Sunkin and Stringer, *Mol. Microbiol.* 19:283-295, 1996; Wada and Nakamura, *DNA Res.* 3:55-64, 1996; Wada *et al., J. Infect. Dis.* 171:1563-1568, 1995). These studies suggest that *P. carinii* has developed an elaborate system for antigenic variation, presumably to evade host defense mechanisms.

Molecular and immunological studies have clearly demonstrated that *P. carinii* isolated from different host species are distinct organisms, and may in fact be separate species (Gigliotti, *J. Infect. Dis.* 165:329-336, 1992; Keely *et al., J. Eukaryot. Microbiol.* 41:94S, 1994; Kovacs *et al., J. Infect. Dis.* 159:60-70, 1989; Stringer, *Infect. Agents Dis.* 2:109-117, 1993). There is a high level of variation among orthologous genes, including the *MSG* genes, isolated from different host-specific strains of the *Pneumocystis*. Hence, diagnosis of *P. carinii* infection in human patients ideally requires *P. carinii* sp. f. hominis (hereinafter "human-*P. carinii*") derived reagents.

The cloning of human-*P. carinii MSG* genes has recently been reported (Garbe and Stringer, *Infect. Immun.* 62:3092-3101, 1994, 1994; Stringer *et al., J. Eukaryot. Microbiol.* 40:821-826, 1993); however, only one full-length sequence was reported.

SUMMARY OF THE INVENTION

The inventors have discovered that human-*P. carinii* MSG is encoded for by a large, highly-conserved gene family, with a particularly conserved region of about 100 amino acids in the C-terminal region of the proteins. They have further discovered that direct detection or nucleic acid amplification (*e.g.*, PCR amplification) of human-*P. carinii* MSG-encoding genes provides a particularly sensitive and specific technique for the detection of *P. carinii*, and the diagnosis of PCP.

This invention encompasses the purified novel human-*P. carinii* proteins represented by SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, and SEQ ID NO: 14, and isolated nucleic acid molecules that encode these proteins. Specific nucleic acid molecules encompassed in this invention include those represented in SEQ ID NO: 1; SEQ ID NO: 2; SEQ ID NO: 3; SEQ ID NO: 4, SEQ ID NO: 5; SEQ ID NO: 6, SEQ ID NO: 7; SEQ ID NO: 15; and SEQ ID NO: 17. Also encompassed within this invention are the isolated nucleic acid sequences that encode the carboxy-terminal conserved about 100 amino acids of the disclosed human-*P. carinii MSG*s; these may be used for amplification or as probes. The sequences of these conserved nucleic acid molecule regions include residues 2794-3042 of *HMSGp1* (SEQ ID NO: 1), 2758-3006 of *HMSGp3* (SEQ ID NO: 3), 2845-3090 of *HMSG11* (SEQ ID NO: 5), 2839-3084 of *HMSG14* (SEQ ID NO: 7), 2836-3081 of *HMSG32* (SEQ ID NO: 9), 2809-3054 of *HMSG33* (SEQ ID NO: 11), 2821-3072 of *HMSG35* (SEQ ID NO: 13), or 1-249 of *HMSGp2* (SEQ ID NO: 15). In addition, this invention encompasses sequences with at least 70% sequence identity to these regions, and recombinant vectors comprising such nucleic acid molecules and conserved regions from within such nucleic acid molecules, as well as transgenic cells including such a recombinant vector.

Another aspect of this invention provides a method of detecting the presence of *Pneumocystis carinii* in a biological specimen, by amplifying with a nucleic acid amplification method (*e.g.*, the polymerase chain reaction) a human-*P. carinii* nucleic acid sequence using two or more oligonucleotide primers derived from a human-*P. carinii* MSG protein encoding sequence, then determining whether an amplified sequence is present. In a preferred embodiment of this invention, the human-*P. carinii* nucleic acid sequence is a highly conserved region within an MSG-protein encoding sequence. Such a highly conserved region may, for instance, include residues 2794-3042 of *HMSGp1* (SEQ ID NO: 1), 2758-3006 of *HMSGp3* (SEQ ID NO: 3), 2845-3090 of *HMSG11* (SEQ ID NO: 5), 2839-3084 of *HMSG14* (SEQ ID NO: 7), 2836-3081 of *HMSG32* (SEQ ID NO: 9), 2809-3054 of *HMSG33* (SEQ ID NO: 11), 2821-3072 of *HMSG35* (SEQ ID NO: 13), or 1-249 of *HMSGp2* (SEQ ID NO: 15). A further aspect of this invention is the method of detecting the presence of *Pneumocystis carinii* in a

biological specimen, by determining whether an amplified sequence is present, for instance by electrophoresis and staining of the amplified sequence, or hybridization to a labeled probe of the amplified sequence. Appropriate labels for the hybridization probe include a fluorescent molecule, a chemiluminescent molecule, an enzyme, a co-factor, an enzyme substrate, or a hapten. The nucleotide sequence of such a probe can be chosen from any *MSG* gene sequence that is amplified in the detection method, and for instance can include a nucleic acid sequence according to SEQ ID NO: 19.

Another aspect of this invention is a method of detecting the presence of *Pneumocystis carinii* in a biological specimen by exposing the biological specimen to a probe that hybridizes to a human-*P. carinii* nucleic acid sequence derived from a human-*P. carinii* MSG protein encoding sequence. The labeled probe to be used in this method may, for instance, include the nucleic acid sequence of SEQ ID NO: 19.

This invention also encompasses one or more oligonucleotide primers including at least 15, or at least 20, 25, 30, 35, 40, 50, or 100, contiguous nucleotides from any of the highly conserved regions within an MSG protein encoding sequence disclosed herein, or from any nucleic acid sequences having at least 70%, or at least 90% or 95%, sequence homology with these sequences. Specific examples of such oligonucleotide primer sequences are shown in SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 23, and SEQ ID NO: 24. Of these primers, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, and SEQ ID NO: 23 may serve as upstream primers, while SEQ ID NO: 20 and SEQ ID NO: 24 may serve as downstream primers.

Kits for detection of a human-*P. carinii* nucleic acid sequence are another aspect of this invention. Such kits may include at least a pair of primers each comprising at least 15, or at least 20, 25, 30, 35, 40, 45, 50, or 100 contiguous nucleotides of any of the conserved regions of the herein disclosed MSG-encoding sequences, and homologs having at least 70% identity with such sequences. Representative primers include those represented by the nucleotide sequences of SEQ ID NO: 17; SEQ ID NO: 18; SEQ ID NO: 19; SEQ ID NO: 20; SEQ ID NO: 21; SEQ ID NO: 22; SEQ ID NO: 23; and SEQ ID NO: 24. These kits may further include a positive nucleic acid amplification (*e.g.*, PCR) control sequence.

Antibodies raised to the peptide sequence according to SEQ ID NO: 25 or SEQ ID NO: 26 are also included within the scope of this invention.

The foregoing and other objects, features, and advantages of the invention will become more apparent from the following detailed description of several embodiments, which proceeds with reference to the accompanying figure and tables.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A-1M is an alignment of the deduced amino acid sequences encoded by two of the human -*P. carinii MSG* genes contained in the genomic clone (*HMSGp1*, SEQ ID NO: 2; and *HMSGp3*, SEQ ID NO: 4) and the five genes generated by PCR (*HMSG11*, SEQ ID NO: 6; *HMSG14*, SEQ ID NO: 8; *HMSG32*, SEQ ID NO: 10; *HMSG33*, SEQ ID NO: 12 and *HMSG35*, SEQ ID NO: 14), together with a published sequence (*GBHMSG*) and a rat-*P. carinii MSG* sequence (*RMSGGP3*, GenBank Accession No: L05906). A methionine was substituted for valine at position 1 in the PCR clones during amplification to facilitate expression, and thus is excluded from the alignment. The peptides that were synthesized and used to generate anti-peptide antibodies are shaded in Figure 1L in light grey (conserved epitope) or dark grey (HMSG32-specific epitope). The arrows (Figure 1L) flank the conserved region that was expressed in pET28a. The conserved carboxy-terminal region of the proteins is boxed (Figure 1L).

SEQUENCE LISTING

The nucleic and amino acid sequences listed in the accompanying sequence listing are shown using standard letter abbreviations for nucleotide bases, and three letter code for amino acids. Only one strand of each nucleic acid sequence is shown, but the complementary strand is understood as included by any reference to the displayed strand.

SEQ ID NO: 1 shows the nucleic acid sequence of *MSG HMSGp1*, GenBank Accession No: AF038556.

SEQ ID NO: 2 shows the amino acid sequence of MSG protein HMSGp1.

SEQ ID NO: 3 shows the nucleic acid sequence of *MSG HMSGp3*, GenBank Accession No: AF038556.

SEQ ID NO: 4 shows the amino acid sequence of MSG protein HMSGp3.

SEQ ID NO: 5 shows the nucleic acid sequence of *MSG HMSG11*, GenBank Accession No: AF033208.

SEQ ID NO: 6 shows the amino acid sequence of MSG protein HuMSG11.

SEQ ID NO: 7 shows the nucleic acid sequence of *MSG HMSG14*, GenBank Accession No: AF033209.

SEQ ID NO: 8 shows the amino acid sequence of MSG protein HuMSG14.

SEQ ID NO: 9 shows the nucleic acid sequence of *MSG HMSG32*, GenBank Accession No: AF033212.

SEQ ID NO: 10 shows the amino acid sequence of MSG protein HuMSG32.

SEQ ID NO: 11 shows the nucleic acid sequence of *MSG HMSG33*, GenBank Accession No: AF033210.

SEQ ID NO: 12 shows the amino acid sequence of MSG protein HuMSG33.

SEQ ID NO: 13 shows the nucleic acid sequence of *MSG HMSG35*, GenBank Accession No: AF033211.

SEQ ID NO: 14 shows the amino acid sequence of MSG protein HMSG35.

SEQ ID NO: 15 shows the nucleic acid sequence of the conserved carboxy-terminal portion of *MSG HMSGp2*, GenBank Accession Number: AF038556.

SEQ ID NO: 16 shows the amino acid sequence of the conserved carboxy-terminal portion of MSG protein HMSGp2.

SEQ ID NO: 17 shows oligonucleotide JKK14 (upstream primer).

SEQ ID NO: 18 shows oligonucleotide JKK15 (upstream primer).

SEQ ID NO: 19 shows oligonucleotide JKK16 (internal probe).

SEQ ID NO: 20 shows oligonucleotide JKK17 (downstream primer).

SEQ ID NO: 21 shows oligonucleotide JK151 (upstream cloning primer).

SEQ ID NO: 22 shows oligonucleotide JK152 (downstream cloning primer).

SEQ ID NO: 23 shows oligonucleotide JK451 (upstream C-terminal cloning primer).

SEQ ID NO: 24 shows oligonucleotide JK452 (downstream C-terminal cloning primer).

SEQ ID NO: 25 shows the amino acid sequence of the internal peptide used to generate antibodies.

SEQ ID NO: 26 shows the amino acid sequence of the C-terminal peptide used to generate antibodies.

DETAILED DESCRIPTION OF THE INVENTION

I. Abbreviations and Definitions

A. Abbreviations

PCP: *Pneumocystis carinii* pneumonia (pneumocystosis)

MSG: major surface glycoprotein

human-*P. carinii*: *P. carinii* sp. f. hominis, human-derived *Pneumocystis carinii*

B. Definitions

Unless otherwise noted, technical terms are used according to conventional usage. Definitions of common terms in molecular biology may be found in Benjamin Lewin, *Genes V*, published by Oxford University Press, 1994 (ISBN 0-19-854287-9); Kendrew *et al.* (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9); and Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8).

In order to facilitate review of the various embodiments of the invention, the following definitions of terms are provided:

Biological Specimen: A biological specimen is a sample of bodily fluid or tissue used for laboratory testing or examination. As used herein, biological specimens include all clinical samples useful for detection of microbial infection in subjects.

Appropriate tissue samples may be taken from the oropharyngeal tract, for instance from lung or bronchial tissue. Samples can be taken by biopsy or during autopsy examination, as appropriate. Biological fluids include blood, derivatives and fractions of blood such as serum, and fluids of the oropharyngeal tract, such as sputum.

Examples of appropriate specimens for use with the current invention for the detection of *P. carinii* include conventional clinical samples, for instance blood or blood-

fractions (*e.g.*, serum), and bronchoalveolar lavage (BAL), sputum, and induced sputum samples. Techniques for acquisition of such samples are well known in the art. Blood and blood fractions (*e.g.*, serum) can be prepared in traditional ways. Oropharyngeal tract fluids can be acquired through conventional techniques, including sputum induction, bronchoalveolar lavage (BAL), and oral washing. Oral washing provides an excellent, non-invasive technique for acquiring appropriate samples to be used in nucleic acid amplification (*e.g.*, PCR) of human-*P. carinii* MSG sequences. Obtaining a sample from oral washing involves having the subject gargle with an amount normal saline for about 10-30 seconds and then expectorate the wash into a sample cup.

cDNA (complementary DNA): A piece of DNA lacking internal, non-coding segments (introns) and transcriptional regulatory sequences. cDNA may also contain untranslated regions (UTRs) that are responsible for translational control in the corresponding RNA molecule. cDNA is synthesized in the laboratory by reverse transcription from messenger RNA extracted from cells.

Isolated: An "isolated" biological component (such as a nucleic acid molecule, protein or organelle) has been substantially separated or purified away from other biological components in the cell of the organism in which the component naturally occurs, *i.e.*, other chromosomal and extra-chromosomal DNA and RNA, proteins and organelles. Nucleic acids and proteins that have been "isolated" include nucleic acids and proteins purified by standard purification methods. The term also embraces nucleic acids and proteins prepared by recombinant expression in a host cell as well as chemically synthesized nucleic acids.

Oligonucleotide: A linear polynucleotide sequence of between 10 and 100 nucleotide bases in length.

Operably linked: A first nucleic acid sequence is operably linked with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences are contiguous and, where necessary to join two protein-coding regions, in the same reading frame.

ORF (open reading frame): A series of nucleotide triplets (codons) coding for amino acids without any internal termination codons. These sequences are usually translatable into a peptide.

Ortholog: Two nucleic acid or amino acid sequences are orthologs of each other if they share a common ancestral sequence and diverged when a species carrying that ancestral sequence split into two species. *P. carinii* isolated from different host species (for instance rats and humans) are known to be distinct organisms, and may in fact be separate *Pneumocystis* species. Because of this, genes and proteins derived from *P. carinii* isolated from different host species are orthologous to each other (*e.g.*, the *MSG11* gene isolated from human-*P. carinii* (*HMSG11*) would be an ortholog of *MSG11* isolated from rat-*P. carinii*). Orthologous sequences are also homologous sequences.

Probes and primers: Nucleic acid probes and primers can be readily prepared based on the nucleic acid molecules provided in this invention. A probe comprises an isolated nucleic acid attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, enzyme substrates, co-factors, ligands, chemiluminescent or fluorescent agents, haptens, and enzymes. Methods for labeling and guidance in the choice of labels appropriate for various purposes are discussed, *e.g.*, in Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York, 1989) and Ausubel *et al.* (In *Current Protocols in Molecular Biology*, Greene Publ. Assoc. and Wiley-Intersciences, 1992).

Primers are short nucleic acid molecules, preferably DNA oligonucleotides 15 nucleotides or more in length. Primers can be annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, and then the primer extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, *e.g.*, by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods known in the art.

Methods for preparing and using probes and primers are described, for example, in Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York, 1989), Ausubel *et al.* (In *Current Protocols in Molecular Biology*, Greene Publ. Assoc. and Wiley-Intersciences, 1992), and Innis *et al.* (In *PCR Protocols, A Guide to*

Methods and Applications, Academic Press, Inc., San Diego, CA, 1990). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, © 1991, Whitehead Institute for Biomedical Research, Cambridge, MA). One of ordinary skill in the art will appreciate that the specificity of a particular probe or primer increases with its length. Thus, for example, a primer comprising 20 consecutive nucleotides of the human-*P. carinii MSG11* gene will anneal to a target sequence, such as another *MSG* gene homolog from the gene family contained within a human-*P. carinii* genomic DNA library, with a higher specificity than a corresponding primer of only 15 nucleotides. Thus, in order to obtain greater specificity, probes and primers can be selected that comprise 20, 25, 30, 35, 40, 50 or more consecutive nucleotides of human-*P. carinii MSG* gene sequences.

The invention thus includes isolated nucleic acid molecules that comprise specified lengths of the disclosed human-*P. carinii MSG* gene sequences. Such molecules may comprise at least 20, 25, 30, 35, 40 or 50 consecutive nucleotides of these sequences, and may be obtained from any region of the disclosed sequences. By way of example, the human-*P. carinii MSG* gene sequences may be apportioned into halves or quarters based on sequence length, and the isolated nucleic acid molecules may be derived from the first or second halves of the molecules, or any of the four quarters. The human-*P. carinii MSG11* gene, shown in SEQ ID NO: 3, can be used to illustrate this. The human-*P. carinii MSG11* gene is 3088 nucleotides in length and so may be hypothetically divided into about halves (nucleotides 1-1544 and 1545-3088) or about quarters (nucleotides 1-772, 773-1544, 1545-2371 and 2372-3088), for instance. Nucleic acid molecules may be selected that comprise at least 20, 25, 30, 35, 40 or 50 consecutive nucleotides of any of these portions of the human-*P. carinii MSG11* gene. Thus, one such nucleic acid molecule might comprise at least 25 consecutive nucleotides of the region comprising nucleotides 2372-3088 of the disclosed human-*P. carinii MSG11* gene (SEQ ID NO: 5).

Further nucleic acid molecules might comprise at least 15 consecutive nucleotides of the regions encoding the conserved carboxy-terminal portion of each human-*P. carinii MSG* gene. These regions comprise nucleotides 2794-3042 of *HMSGp1* (SEQ ID NO: 1), 2758-3006 of *HMSGp3* (SEQ ID NO: 3), 2845-3090 of *HMSG11* (SEQ ID NO: 5), 2839-3084 of *HMSG14* (SEQ ID NO: 7), 2836-3081 of *HMSG32* (SEQ ID NO: 9), 2809-3054

of *HMSG33* (SEQ ID NO: 11), 2821-3072 of *HMSG35* (SEQ ID NO: 13), and 1-249 of *HMSGp2* (SEQ ID NO: 15), respectively.

Recombinant: A recombinant nucleic acid is one that has a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two otherwise separated segments of sequence. This artificial combination can be accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, *e.g.*, by genetic engineering techniques.

Sequence identity: The similarity between two nucleic acid sequences, or two amino acid sequences, is expressed in terms of the similarity between the sequences, otherwise referred to as sequence identity. Sequence identity is frequently measured in terms of percentage identity (or similarity or homology); the higher the percentage, the more similar the two sequences are. Homologs of human-*P. carinii* MSG proteins, and the corresponding gene sequences, will possess a relatively high degree of sequence identity when aligned using standard methods. This homology will be more significant when the proteins or gene sequences are derived from *P. carinii* isolated from one host species (*i.e.*, two human-*P. carinii* MSG homologs will typically have greater sequence identity than that shown by one human- and one rat-*P. carinii* MSG ortholog).

Typically, human-*P. carinii* MSG homologs are 74 to 91% identical at the nucleotide level and 63 to 88% identical at the amino acid level when comparing pairs of clones. In comparison, there is approximately 60% identity at the DNA level and 40% identity at the amino acid level when comparing a human-P. *carinii* MSG to the rat-*P. carinii* ortholog MSGGP3.

Methods of alignment of sequences for comparison are well known in the art. Various programs and alignment algorithms are described in: Smith & Waterman, *Adv. Appl. Math.* 2:482, 1981; Needleman & Wunsch, *J. Mol. Biol.* 48:443, 1970; Pearson & Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444, 1988; Higgins & Sharp, *Gene* 73:237-244, 1988; Higgins & Sharp, *CABIOS* 5:151-153, 1989; Corpet *et al., Nuc. Acids Res.* 16:10881-10890, 1988; Huang *et al., Computer Appls. in the Biosciences* 8:155-165, 1992; and Pearson *et al., Meth. Mol. Bio.* 24:307-331, 1994. Altschul *et al., J. Mol. Biol.* 215:403-410, 1990, presents a detailed consideration of sequence alignment methods and homology calculations.

The NCBI Basic Local Alignment Search Tool (BLAST) (Altschul *et al., J. Mol. Biol.* 215:403-410, 1990) is available from several sources, including the National Center for Biotechnology Information (NCBI, Bethesda, MD) and on the Internet, for use in connection with the sequence analysis programs blastp, blastn, blastx, tblastn and tblastx. It can be accessed at the NCBI online site under the "BLAST" heading. A description of how to determine sequence identity using this program is available at the NCBI online site under the "BLAST" heading and "BLAST overview" subheading. For comparisons of amino acid sequences of greater than about 30 amino acids, the Blast 2.0 sequences function is employed using the default BLOSUM62 matrix set to default parameters, (gap existence cost of 11, and a per residue gap cost of 1). When aligning short peptides (fewer than around 30 amino acids), the alignment should be performed using the Blast 2.0 sequences function, employing the PAM30 matrix set to default parameters (open gap 9, extension gap 1 penalties).

Other members of the gene family of the disclosed human-*P. carinii* MSG proteins typically possess at least 60% sequence identity counted over full-length alignment with the amino acid sequence of human-*P. carinii* MSG using the NCBI Blast 2.0, gapped blastp set to default parameters. Sequence identity over the about 100 C-terminal amino acids will typically be higher than 60%, for instance about 63%. Proteins with even greater similarity to the reference sequence will show increasing percentage identities when assessed by this method, such as at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, or at least 98% sequence identity. When less than the entire sequence is being compared for sequence identity, homologs will typically possess at least 75% sequence identity over short windows of 10-20 amino acids, and may possess sequence identities of at least 85% or at least 90% or at least 95% depending on their similarity to the reference sequence. Methods for determining sequence identity over such short windows are described at the NCBI online site under the "BLAST" heading and "Frequently Asked Questions" subheading.

One of ordinary skill in the art will appreciate that these sequence identity ranges are provided for guidance only; it is entirely possible that strongly significant homologs could be obtained that fall outside of the ranges provided. The present invention provides

not only the peptide homologs that are described above, but also nucleic acid molecules that encode such homologs.

An alternative indication that two nucleic acid molecules are closely related is that the two molecules hybridize to each other under stringent conditions. Stringent conditions are sequence-dependent and are different under different environmental parameters. Generally, stringent conditions are selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Conditions for nucleic acid hybridization and calculation of stringencies can be found in Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York, 1989) and Tijssen (In *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes* Part I, Chapter 2, Elsevier, New York, 1993). Nucleic acid molecules that hybridize under stringent conditions to a human-*P. carinii MSG* gene sequence will typically hybridize to a probe based on either an entire human-*P. carinii MSG* gene or selected portions of the gene under wash conditions of 2 x SSC at 50°C. A more detailed discussion of hybridization conditions is presented below.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences, due to the degeneracy of the genetic code. It is understood that changes in nucleic acid sequences can be made using this degeneracy to produce multiple nucleic acid molecules that all encode substantially the same protein.

Specific binding agent: An agent that binds substantially only to a defined target. Thus an MSG protein-specific binding agent binds substantially only the MSG protein. As used herein, the term "MSG protein specific binding agent" includes anti-MSG protein antibodies and other agents that bind substantially only to the MSG protein.

Anti-MSG protein antibodies may be produced using standard procedures described in a number of texts, including Harlow and Lane (*Antibodies, A Laboratory Manual*, CSHL, New York, 1988). The determination that a particular agent binds substantially only to the MSG protein may readily be made by using or adapting routine procedures. One suitable *in vitro* assay makes use of the Western blotting procedure

(described in many standard texts, including Harlow and Lane (In *Antibodies, A Laboratory Manual*, CSHL, New York, 1988)). Western blotting may be used to determine that a given MSG protein binding agent, such as an anti-MSG protein monoclonal antibody, binds substantially only to the MSG protein.

Shorter fragments of antibodies can also serve as specific binding agents. For instance, FAbs, Fvs, and single-chain Fvs (SCFvs) that bind to MSG would be MSG-specific binding agents.

Transformed: A transformed cell is a cell into which has been introduced a nucleic acid molecule by molecular biology techniques. As used herein, the term transformation encompasses all techniques by which a nucleic acid molecule might be introduced into such a cell, including transfection with viral vectors, transformation with plasmid vectors, and introduction of naked DNA by electroporation, lipofection, and particle gun acceleration.

Vector: A nucleic acid molecule as introduced into a host cell, thereby producing a transformed host cell. A vector may include nucleic acid sequences that permit it to replicate in a host cell, such as an origin of replication. A vector may also include one or more selectable marker genes and other genetic elements known in the art.

II. Human-*P. Carinii MSG* Sequences

This specification provides MSG proteins and MSG-encoding nucleic acid molecules, including gene sequences, derived from human-*P. carinii*. The prototypical *MSG* sequences are the human-*P. carinii* sequences as presented herein (*HMSGp1*, *HMSGp3*, *HMSG11*, *HMSG14*, *HMSG32*, *HMSG33*, and *HMSG* 35).

a. Human-*P. carinii HMSGp1*, *HMSGp3*, *HMSG11*, *HMSG14*, *HMSG32*, *HMSG33*, and *HMSG35*

Human-*P. carinii HMSGp1*, *HMSGp3*, *HMSG11*, *HMSG14*, *HMSG32*, *HMSG33*, and *HMSG35* genomic sequences are shown in SEQ ID NOS: 1, 3, 5, 7, 9, 11, and 13, respectively. The sequences typically encode proteins that are about 1000 to about 1030 amino acids in length (for instance, SEQ ID NO: 5 shows the amino acid sequence of the MSG11 protein, which is 1028 amino acids long). These human-*P. carinii* MSG proteins show significant sequence similarity to each other, and a lesser degree of sequence similarity to MSG proteins derived from organisms in other hosts.

With the provision herein of seven novel human-*P. carinii MSG* gene sequences, nucleotide amplification methods, for instance polymerase chain reaction (PCR), may now be utilized as a preferred method for producing nucleic acid sequences encoding these human-*P. carinii* MSG proteins. For example, PCR amplification of the human-*P. carinii MSG11* gene sequence may be accomplished by direct PCR from a clinical sample. Methods and conditions for direct PCR are known in the art and are described in Innis *et al.* (*PCR Protocols, A Guide to Methods and Applications*, Academic Press, Inc., San Diego, CA, 1990). Appropriate sampling methods are described more fully below.

The selection of amplification primers will be made according to the portions of the gene that are to be amplified. Primers may be chosen to amplify small segments of the gene, the open reading frame, or the entire gene sequence. Variations in amplification conditions may be required to accommodate primers of differing lengths; such considerations are well known in the art and are discussed in Innis *et al.* (*PCR Protocols, A Guide to Methods and Applications*, Academic Press, Inc., San Diego, CA, 1990), Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York, 1989), and Ausubel *et al.* (In *Current Protocols in Molecular Biology*, Greene Publ. Assoc. and Wiley-Intersciences, 1992). By way of example only, the human-*P. carinii HMSG11* gene as shown in SEQ ID NO: 5 can be amplified using the following combination of primers:

primer JK151: 5' TTT CAT ATG GCG CGG GCG GTC AAG CGG CAG 3' (SEQ ID NO: 21)

primer JK152: 5' CTA AAT CAT GAA CGA AAT AAC CAT TGC TAC 3' (SEQ ID NO: 22).

The sequence encoding the conserved carboxy-terminal region of human-*P. carinii HMSG11* can be amplified using the following primer pair:

primer JKK14: 5' GAA TGC AAA TCC TTA CAG ACA ACA G 3' (SEQ ID NO: 17)

primer JKK17: 5' AAA TCA TGA ACG AAA TAA CCA TTG C 3' (SEQ ID NO: 20).

These primers are illustrative only; one skilled in the art will appreciate that many different primers may be derived from the provided *MSG* gene sequences in order to amplify particular regions of these molecules. Resequencing of PCR products obtained by these amplification procedures is recommended; this will facilitate confirmation of the amplified sequence and will also provide information on natural variation on this sequence in different ecotypes and plant populations. Oligonucleotides derived from the human-*P. carinii MSG* gene sequences provided may be used in such sequencing methods.

Further homologous human-*P. carinii MSGs* can be cloned in a similar manner. In order to increase the number of *MSG*s that can be amplified in a single PCR reaction, a third primer can be added. For instance, a second upstream primer (*e.g.*, primer JKK15: 5' GAA TGC AAA TCT TTA CAG ACA ACA G 3' (SEQ ID NO: 18)) may be added to the amplification reaction along with primers JKK14 and JKK17. Typically, when more than two primers are provided in a single PCR amplification reaction, those primers that anneal to the same site on the target nucleotide sequence (*e.g.*, JKK14 and JKK15) will be provided in equimolar amounts (for instance, 0.625 pM each), and such that the total amount of primer provided for each end of the amplicon will be equivalent (for instance, 1.25 pM each).

Oligonucleotides that are derived from the human-*P. carinii HMSGp1*, *HMSGp3*, *HMSG11*, *HMSG14*, *HMSG32*, *HMSG33*, and *HMSG35* gene sequences (SEQ ID NOS: 1, 3, 5, 7, 9, 11, and 13, respectively), as well as the fragment of *HMSGp2* disclosed (SEQ ID NO: 15), are encompassed within the scope of the present invention. Preferably, such oligonucleotide primers will comprise a sequence of at least 15-20 consecutive nucleotides of the relevant human-*P. carinii MSG* gene sequence. To enhance amplification specificity, oligonucleotide primers comprising at least 25, 30, 35, 40, 45 or 50 consecutive nucleotides of these sequences may also be used. These primers for instance may be obtained from any region of the disclosed sequences. By way of example, human-*P. carinii MSG* gene sequences may be apportioned into halves

or quarters based on sequence length, and the isolated nucleic acid molecules may be derived from the first or second halves of the molecules, or any of the four quarters. In addition, primers may be specifically chosen from the conserved carboxy-terminal region of each *MSG* coding sequence. This region comprises nucleic acid residues 2794-3042 of *HMSGp1* (SEQ ID NO: 1), 2758-3006 of *HMSGp3* (SEQ ID NO: 3), 2845-3090 of *HMSG11* (SEQ ID NO: 5), 2839-3084 of *HMSG14* (SEQ ID NO: 7), 2836-3081 of *HMSG32* (SEQ ID NO: 9), 2809-3054 of *HMSG33* (SEQ ID NO: 11), 2821-3072 of *HMSG35* (SEQ ID NO: 13), and 1-249 of *HMSGp2* (SEQ ID NO: 15).

b. MSG Sequence Variants

With the provision of human-*P. carinii* HMSGp1, HMSGp3, HMSG11, HMSG14, HMSG32, HMSG33, and HMSG35 proteins and corresponding gene sequences herein, the creation of variants of these sequences is now enabled.

Variant MSG proteins include proteins that differ in amino acid sequence from the human-*P. carinii* MSG sequences disclosed but that share at least 63% amino acid sequence homology (for example at least 80%, 90%, 95% or 98% homology) with any of the provided human MSG proteins. Such variants may be produced by manipulating the nucleotide sequence of the, for instance, human-*P. carinii HMSG11* gene using standard procedures, including for instance site-directed mutagenesis or PCR. The simplest modifications involve the substitution of one or more amino acids for amino acids having similar biochemical properties. These so-called conservative substitutions are likely to have minimal impact on the activity of the resultant protein. Table 1 shows amino acids that may be substituted for an original amino acid in a protein, and which are regarded as conservative substitutions.

Table 1.

Original Residue	Conservative Substitutions
Ala	ser
Arg	lys
Asn	gln; his
Asp	glu
Cys	ser
Gln	asn
Glu	asp
Gly	pro
His	asn; gln

Original Residue	Conservative Substitutions
Ile	leu; val
Leu	ile; val
Lys	arg; gln; glu
Met	leu; ile
Phe	met; leu; tyr
Ser	thr
Thr	ser
Trp	tyr
Tyr	trp; phe
Val	ile; leu

More substantial changes in enzymatic function or other protein features may be obtained by selecting amino acid substitutions that are less conservative than those listed in Table 1. Such changes include changing residues that differ more significantly in their effect on maintaining polypeptide backbone structure (*e.g.*, sheet or helical conformation) near the substitution, charge or hydrophobicity of the molecule at the target site, or bulk of a specific side chain. The following substitutions are generally expected to produce the greatest changes in protein properties: (a) a hydrophilic residue (*e.g.*, seryl or threonyl) is substituted for (or by) a hydrophobic residue (*e.g.*, leucyl, isoleucyl, phenylalanyl, valyl or alanyl); (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain (*e.g.*, lysyl, arginyl, or histadyl) is substituted for (or by) an electronegative residue (*e.g.*, glutamyl or aspartyl); or (d) a residue having a bulky side chain (*e.g.*, phenylalanine) is substituted for (or by) one lacking a side chain (*e.g.*, glycine).

Variant *MSG* genes may be produced by standard DNA mutagenesis techniques, for example, M13 primer mutagenesis. Details of these techniques are provided in Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York, 1989), Ch. 15. By the use of such techniques, variants may be created which differ in minor ways from the human-*P. carinii MSG* gene sequences disclosed. DNA molecules and nucleotide sequences which are derivatives of those specifically disclosed herein and that differ from those disclosed by the deletion, addition, or substitution of nucleotides while still encoding a protein that has at least 63% sequence identity with the *MSG* sequences disclosed (SEQ ID NOS: 1, 3, 5, 7, 9, 11, and 13) are comprehended by this invention. In their most simple form, such variants may differ from the disclosed

sequences by alteration of the coding region to fit the codon usage bias of the particular organism into which the molecule is to be introduced.

Alternatively, the coding region may be altered by taking advantage of the degeneracy of the genetic code to alter the coding sequence such that, while the nucleotide sequence is substantially altered, it nevertheless encodes a protein having an amino acid sequence substantially similar to the disclosed human-*P. carinii* MSG protein sequences. For example, the second amino acid residue of the human-*P. carinii* HMSG11 protein is alanine. The nucleotide codon triplet GCG encodes this alanine residue. Because of the degeneracy of the genetic code, three other nucleotide codon triplets - GCT, GCC and GCA - also code for alanine. Thus, the nucleotide sequence of the human-*P. carinii HMSG11* ORF could be changed at this position to any of these three alternative codons without affecting the amino acid composition or characteristics of the encoded protein. Based upon the degeneracy of the genetic code, variant DNA molecules may be derived from the cDNA and gene sequences disclosed herein using standard DNA mutagenesis techniques as described above, or by synthesis of DNA sequences. Thus, this invention also encompasses nucleic acid sequences which encode an MSG protein, but which vary from the disclosed nucleic acid sequences by virtue of the degeneracy of the genetic code.

Variants of the MSG protein may also be defined in terms of their sequence identity with the prototype MSG proteins shown in SEQ ID NOS: 2, 4, 6, 8, 10, 12, and 14. As described above, human MSG proteins share at least 60% (for example, at least 63%) amino acid sequence identity with the human-*P. carinii* HMSGp1, HMSGp3, HMSG11, HMSG14, HMSG32, HMSG33, or HMSG35 proteins (SEQ ID NOS: 2, 4, 6, 8, 10, 12, and 14, respectively). Nucleic acid sequences that encode such proteins may readily be determined simply by applying the genetic code to the amino acid sequence of an MSG protein, and such nucleic acid molecules may readily be produced by assembling oligonucleotides corresponding to portions of the sequence.

Nucleic acid molecules that are derived from the human-*P. carinii MSG* gene sequences disclosed include molecules that hybridize under stringent conditions to the disclosed prototypical *MSG* nucleic acid molecules, or fragments thereof. Stringent conditions are hybridization at 65°C in 6 x SSC, 5 x Denhardt's solution, 0.5% SDS and

100 μg sheared salmon testes DNA, followed by 15-30 minute sequential washes at 65°C in 2 x SSC, 0.5% SDS, followed by 1 x SSC, 0.5% SDS and finally 0.2 x SSC, 0.5% SDS.

Low stringency hybridization conditions (to detect less closely related homologs) are performed as described above but at 50°C (both hybridization and wash conditions); however, depending on the strength of the detected signal, the wash steps may be terminated after the first 2 x SSC wash.

Human-*P. carinii HMSGp1*, *HMSGp3*, *HMSG11*, *HMSG14*, HMSG32, *HMSG33*, and *HMSG35* genes (SEQ ID NOS: 1, 3, 5, 7, 9, 11 and 13, respectively), as well as the fragment of *HMSGp2* disclosed (SEQ ID NO: 15), and homologs of these sequences may be incorporated into transformation or expression vectors.

III. Detection of *P. Carinii* in Clinical Specimens

The conserved nature of human-*P. carinii MSG* genes provided in this specification, and particularly the highly-conserved about 100 amino acid region in the C-terminal portion of the protein, makes these genes useful targets for use in detection of *P. carinii* in clinical samples and diagnosis of PCP.

a. Clinical Specimens

Appropriate specimens for use with the current invention in detection of *P. carinii* include any conventional clinical samples, for instance blood or blood-fractions (*e.g.*, serum), and bronchoalveolar lavage (BAL), sputum, and induced sputum samples. Techniques for acquisition of such samples are well known in the art. See, for instance, Schluger *et al.* (*J. Exp. Med* 176:1327-1333, 1992) (collection of serum samples); Bigby *et al.* (*Am. Rev. Respir. Dis.* 133:515-518, 1986) and Kovacs *et al.* (*NEJM* 318:589-593, 1988) (collection of sputum samples); and Ognibene *et al.* (*Am. Rev. Respir. Dis.* 129:929-932, 1984) (collection of bronchoalveolar lavage (BAL)).

In addition to conventional methods, oral washing provides an excellent, non-invasive technique for acquiring appropriate samples to be used in nucleic acid amplification (*e.g.*, PCR) of human-*P. carinii MSG* sequences (Helweg-Larsen *et al., J. Clin. Microbiol.* 36:2068-2072, 1998). Oral washing involves having the subject gargle

with 50 cc of normal saline for 10-30 seconds and then expectorate the wash into a sample cup.

Serum or other blood fractions can be prepared in the conventional manner. About 200 μL of serum is an appropriate amount for the extraction of DNA for use in amplification reactions. See also, Schluger *et al.*, *J. Exp. Med.* 176:1327-1333, 1992; Ortona *et al.*, *Mol. Cell Probes* 10:187-190, 1996.

Once a sample has been obtained, DNA can be extracted through any conventional method. For instance, rapid DNA preparation can be performed using a commercially available kit (*e.g.*, the InstaGene Matrix, BioRad, Hercules, CA; the NucliSens isolation kit, Organon Teknika, Netherlands). Preferably the DNA preparation technique chosen yields a nucleotide preparation that is accessible to and amenable to nucleic acid amplification.

b. Direct Hybridization Probing Detection

Human-*P. carinii MSG* gene sequences can be detected through the hybridization of an oligonucleotide probe to nucleic acid molecules prepared from a clinical sample. The sequence of appropriate oligonucleotide probes will correspond to a region within one or more of the human-*P. carinii MSG* sequences disclosed herein. Techniques for use in hybridization of oligonucleotide probes to target sequences will be known to one of ordinary skill in the art. See, for instance, U.S. Patent Nos. 5,164,490 (disclosing use of sequences from the *P. carinii* dihydrofolate reductase gene as direct hybridization probes) and 5,519,127 (using nucleic acid probes capable of hybridizing to rRNA or rDNA of *P. carinii* for detection of the organism). In general, hybridization probes will be at least 15 bases in length, and may be 20, 25, 30, 35, 40 or 50 or more bases in length. For instance, a probe may comprise the entire conserved sequence of an *MSG* (*e.g.*, residues 2845-3090 of *HMSG11*), or the entire coding sequence of the gene. Typically such a probe will be detectably labeled in some fashion, either with an isotopic or non-isotopic label. Such non-isotopic labels may, for instance, comprise a fluorescent or luminescent molecule, or an enzyme, co-factor, enzyme substrate, or hapten. The probe is generally incubated with a single-stranded preparation of DNA, RNA, or a mixture of both, and hybridization determined after separation of double and single-stranded molecules. Alternatively, probes may be incubated with a nucleotide

preparation after it has been separated by size and/or charge and immobilized on an appropriate medium. Hybridization techniques suitable for use with oligonucleotides are well known to those of ordinary skill in the art. For general references on the conditions and options that are appropriate, see Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York, 1989) and Ausubel *et al.* (In *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley-Intersciences, 1992).

c. Nucleic Acid-Mediated Detection

It may be advantageous to amplify target *P. carinii* gene sequences in a clinical sample prior to using a hybridization probe to detect its presence. For instance, for detection of human-*P. carinii MSG* gene sequences, it may be advantageous to amplify part or all of the *MSG* gene sequence, then detect the presence of the amplified sequence pool. Any nucleic acid amplification method can be used, including polymerase chain reaction (PCR) amplification. Amplification can be carried out in a simple single reaction using a pair of primers, or can be enhanced by the use of multiple degenerate primers to increase the number of *MSG* homologs that are amplified. Where degenerate primers are used, the sequence variability of the disclosed human-*P. carinii MSG* gene sequences can be used to design appropriate primers that will be specific for multiple human-*P. carinii MSG* homologs. Alternately, amplification specificity can be increased through the use of nested PCR techniques, which are known (see, for instance, Lipschik *et al., Lancet* 340:203-206, 1992, using nested sets of primers to rRNA in the detection of *Pneumocystis carinii*).

It is also possible to run sequential PCR amplification experiments on samples using different targets in each reaction, such that putative positive samples detected in the first reaction are confirmed by amplification of a second sequence. For instance, it would be possible to analyze clinical samples through PCR amplification of a human-*P. carinii MSG* gene, then to take only those samples that are positive for amplification of *MSG* and test them also for the presence of *P. carinii* rRNA. Such sequential testing of samples will help reduce false positive results due to cross contamination of PCR samples; it is unlikely that a clinical sample will become contaminated with both target sequences.

The selection of PCR primers will be made according to the portions of the gene sequence that are to be amplified. For use in PCR detection of *P. carinii*, it is advantageous to choose primer-annealing sites that are highly conserved across many different members of the human-*P. carinii MSG* gene family. For instance, it is advantageous to choose primer sites from within the regions of human-*P. carinii* sequence displaying greater than 63% sequence identity across the disclosed family members, *e.g.*, that portion of the gene encoding the conserved carboxy-terminal region of the protein. The highly conserved carboxy-terminal regions of the disclosed genes are as follows: residues 2794-3042 of *HMSGp1* (SEQ ID NO: 1), 2758-3006 of *HMSGp3* (SEQ ID NO: 3), 2845-3090 of *HMSG11* (SEQ ID NO: 5), 2839-3084 of *HMSG14* (SEQ ID NO: 7), 2836-3081 of *HMSG32* (SEQ ID NO: 9), 2809-3054 of *HMSG33* (SEQ ID NO: 11), 2821-3072 of *HMSG35* (SEQ ID NO: 13), and 1-249 of *HMSGp2* (SEQ ID NO: 15).

Variations in amplification conditions may be required to accommodate primers of differing lengths; such considerations are well known in the art and are discussed in Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York, 1989) and Ausubel *et al.* (In *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley-Intersciences, 1992). By way of example only, primers JKK14, JKK15, and JKK17 (SEQ ID NOS: 17, 18, and 20, respectively) can be used to amplify the C-terminal conserved region of several human-*P. carinii MSG* genes. These primers are illustrative only; one skilled in the art will appreciate that many different primers may be derived from the provided cDNA and gene sequences in order to amplify particular regions of these molecules.

Oligonucleotides to be used in detection of the *P. carinii* organism or diagnosis of PCP that are derived from the human-*P. carinii MSG* gene sequences disclosed herein are encompassed within the scope of the present invention.

d. Detection of Amplified *P. carinii MSG* Sequences

The presence of amplified human-*P. carinii MSG* sequences can be determined in any conventional manner, including electrophoresis and staining (for instance, with ethidium bromide) of the amplified sequence, or hybridization of a labeled probe to the amplified sequence. For general guidelines on such techniques, see Sambrook *et al.*

(*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York, 1989), and Ausubel *et al.* (*Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley-Interscience, 1987). Hybridization probes appropriate for use in detection of amplified human-*P. carinii MSG* sequences are essentially equivalent to those described above for direct hybridization. The region of the gene that has been amplified will be important in choosing an appropriate probe; the detection probe should hybridize to a sequence that falls between the ends of the amplification primers such that the annealing site of the probe is amplified. By way of example, one appropriate oligonucleotide probe is JKK16 (SEQ ID NO: 19), which corresponds to residues of 2926-2950 of *HMSG33*. This probe could be used for detection of both full-length and carboxy-terminal amplified fragments of human-*P. carinii MSG* genes.

Typically, oligonucleotide probes will be labeled as discussed above, and detection will be carried out through conventional methods. In general, detection of amplified sequences will be more sensitive than direct hybridization.

In addition to radioisotope labeled hybridizing probes, amplicons can be detected using fluorescent labeled probes. One such appropriate fluorescent label is europium (Eu^{3+}). See, for instance, Lopez *et al., Clin. Chem.* 39(2):196-201, 1993 (using a europium derivative for time-resolved fluorescence detection of amplified human papillomavirus sequences); Eskola *et al., Clin. Biochem.* 27(5):373-379, 1994 (using PCR and europium-labeled DNA probes to detect a marker for chronic myelogenous leukemia); and Dahlen *et al., J. Clin. Microbiol.* 29(4):798-804, 1991 (detection of PCR amplified HIV sequences using biotinylated and europium labeled oligonucleotide probes).

e. Preparation of a Positive Nucleic Acid Amplification Control

It is advantageous to provide a positive control sequence for use in nucleic acid amplification reactions, to ensure that the system is functioning properly. The positive control sequence should be one the provided oligonucleotide primers are known to anneal to. Therefore, in the present invention, appropriate positive control sequences include, for instance, any sequences that can be amplified with the same primers as are used to amplify human-*P. carinii MSG*. For instance, primers JKK14 (SEQ ID NO: 17) and JKK17 (SEQ ID NO: 20) can serve as appropriate primers. It is advantageous,

however, if the internal amplified sequence is distinguishable from the *MSG* target (*i.e.*, is a mimic rather than identical sequence); this allows specific and separate detection of the target and mimic amplified products. Appropriate differences between the two sequences include overall length of the amplicon (where detection of the PCR products will be performed using electrophoresis and subsequent staining) and amplicon sequence differences (where detection of the PCR products will be performed using hybridization to a labeled probe specific for each amplified sequence).

Nucleic acid amplification positive control sequences can be provided in the form of independent, linear nucleotide sequences. Alternately, a recombinant vector comprising the appropriate positive control sequence may be provided. Construction of such a recombinant vector is by conventional means, and any of a myriad of conventional cloning vectors can be used. In general, the vector will include one or more restriction enzyme sites into which the PCR control sequence can be inserted. The vector may also comprise a replication site to provide for its production in a suitable host cell, for instance in a bacterial cell. The choice of appropriate cloning vector will be within the skill of an ordinary artisan.

IV. Kits for Detection of *P. carinii*

The oligonucleotide primers disclosed herein can be supplied in the form of a kit for use in detection of *P. carinii* or diagnosis of PCP. In such a kit, an appropriate amount of one or more of the oligonucleotide primers is provided in one or more containers. The oligonucleotide primers may be provided suspended in an aqueous solution or as a freeze-dried or lyophilized powder, for instance. The container(s) in which the oligonucleotide(s) are supplied can be any conventional container that is capable of holding the supplied form, for instance, microfuge tubes, ampoules, or bottles. In some applications, pairs of primers may be provided in pre-measured single use amounts in individual, typically disposable, tubes or equivalent containers. With such an arrangement, the sample to be tested for the presence of human-*P. carinii* can be added to the individual tubes and amplification carried out directly.

The amount of each oligonucleotide primer supplied in the kit can be any appropriate amount, depending for instance on the market to which the product is

directed. For instance, if the kit is adapted for research or clinical use, the amount of each oligonucleotide primer provided would likely be an amount sufficient to prime several PCR amplification reactions. Those of ordinary skill in the art know the amount of oligonucleotide primer that is appropriate for use in a single amplification reaction. General guidelines may for instance be found in Innis *et al.* (*PCR Protocols, A Guide to Methods and Applications*, Academic Press, Inc., San Diego, CA, 1990), Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York, 1989), and Ausubel *et al.* (In *Current Protocols in Molecular Biology*, Greene Publ. Assoc. and Wiley-Intersciences, 1992).

A kit may include more than two primers, in order to facilitate the PCR amplification of a larger number of human-*P. carinii MSG* genes. For instance, primers JKK14 (SEQ ID NO: 17) and JKK15 (SEQ ID NO: 18) both may be provided as upstream primers, while primer JKK17 (SEQ ID NO: 20) is provided as a downstream primer. These primers are provided by way of example only.

In some embodiments of the current invention, kits may also include the reagents necessary to carry out PCR amplification reactions, including, for instance, DNA sample preparation reagents, appropriate buffers (*e.g.*, polymerase buffer), salts (*e.g.*, magnesium chloride), and deoxyribonucleotides (dNTPs).

Kits may in addition include either labeled or unlabeled oligonucleotide probes for use in detection of the amplified human-*P. carinii* sequences. The appropriate sequences for such a probe will be any sequence that falls between the annealing sites of the two provided oligonucleotide primers, such that the sequence the probe is complementary to is amplified during the PCR reaction. Primer JKK16 (SEQ ID NO: 19) exemplifies such a sequence, and an appropriate probe could comprise this sequence.

It may also be advantageous to provide in the kit one or more control sequences for use in the PCR reactions. Appropriate positive control sequences may be essentially as those discussed above.

EXAMPLES

Example 1: Isolation of Multiple Human-*P. carinii* *MSG* Sequences.

A. Polymerase Chain Reaction (PCR) Amplification Cloning

DNA was isolated from an autopsy lung sample of an HIV-infected patient with *P. carinii* pneumonia according to standard methods, using SDS and proteinase K (0.5 μg/ml), followed by phenol-chloroform extraction and ethanol precipitation (Davis *et al.*, In *Basic Methods in Molecular Biology*, Elsevier, NY, 1986). A genomic library using the same DNA cloned into the Xho 1 site of lambda GEM 12 vector (Promega, Madison, WI) was commercially prepared (Lofstrand Labs Limited, Gaithersburg, MD).

Primers to amplify full-length human-*P. carinii* genes were designed based on published data (Garbe and Stringer, *Infect. Immun.* 62(8):3092-3101, 1994). The sense primer, JK151 (5'-TTT CAT ATG GCG CGG GCG GTC AAG CGG CAG-3') (SEQ ID NO: 21) corresponds to nucleotides 153 to 175 of a published *MSG* sequence (GenBank Accession No: L27092), and the antisense primer JK152 (5'-CTA AAT CAT GAA CGA AAT AAC CAT TGC TAC-3') (SEQ ID NO: 22) is complementary to nucleotides 3215 to 3244 of the same sequence. An Nde I site was created at the beginning of JK151, which substitutes a methionine for the valine of the original sequence, to facilitate subcloning and expression. For amplification, 1 μg of genomic DNA was added to a 50 μl reaction containing primers (25 pM each), dNTPs (0.2 mM), 5 μl of AmpliTaq (Perkin-Elmer), and $MgCl_2$ (2.5 mM). The DNA amplification was performed on a Perkin Elmer Cetus DNA thermal cycler. An initial denaturation cycle (1 minute at 96°C) was followed by 36 cycles of denaturation at 95°C for 1 minute, annealing at 50°C for 2 minutes and extension at 72°C for 2 minutes, followed by a final extension after the last cycle at 72°C for 10 minutes.

***** A band of the correct size (approximately 3.1 Kb) was amplified and subjected to electrophoresis in 1% agarose gel in 1X TBE buffer. PCR products were then directly subcloned into PCR II (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. Five clones that differed in their restriction mapping and hybridization

patterns were identified and sequenced (*HMSG11* (SEQ ID NO: 5) GenBank Accession No: AF033208; *HMSG14* (SEQ ID NO: 7) GenBank Accession No: AF033209; *HMSG33* (SEQ ID NO: 11) GenBank Accession No: AF033210; *HMSG35* (SEQ ID NO: 13) GenBank Accession No: AF033211; and *HMSG32* (SEQ ID NO: 9) GenBank Accession No: AF033212).

Nucleotide sequencing was performed using an automated sequencer (Model 373 or 377, Applied Biosystems/Perkin Elmer, Foster City, CA). The nucleotide sequence and deduced amino acid sequence data were analyzed by Factura and AutoAssembler (both from Applied Biosystems), Sequencher (Gene Codes Corp., Ann Arbor, MI), MacVector (Scientific Imaging Systems, New Haven, CT), ClustalW (40), and GeneWorks (IntelliGenetics, Mountain View, CA).

All clones encoded *MSG* variants that were clearly related but differed from each other. The coding region of the clones varied in length from 3,054 to 3,087 bases, encoding proteins of 1,008 to 1,028 amino acids with predicted molecular weights of 114 to 117 KDa. They are 74 to 91% identical at the nucleotide level and 63 to 88% identical at the amino acid level when comparing pairs of clones. Overall, approximately 50% of the amino acids are conserved in all five clones. The clones are more closely related to each other than to rat *P. carinii MSG* genes. There is an approximately 60% identity at the DNA level and 40% identity at the amino acid level when comparing a human-*P. carinii MSG* to rat *P. carinii MSGGP3*.

B. Southern hybridization/Library screening

For southern hybridization with a radioactive probe, DNA was treated with restriction enzymes, separated by agarose gel electrophoresis and transferred to Hybond N+ membranes (Amersham, Life Science, Arlington Heights, IL) with 0.4 M NaOH. DNA was probed using an approximately 600 bp Xba I fragment of the human-*P. carinii MSG* III gene (Garbe and Stringer (1994) *Infect. Immuno.* **62**:3092-3101) that had been labeled with α-32P dATP or α-32P dCTP by a random priming kit (Boehringer Mannheim). Filters were prehybridized for 4 hours and then hybridized overnight at 55°C in 6X SSPE with 0.5% SDS, and 5X Denhardt's solution. Blots were washed in 6X

SSPE with 0.5% SDS at room temperature for 10 minutes and then in 0.5X SSPE with 0.5% SDS at 55°C twice for 30 minutes each. The genomic library was screened using a gel-purified full-length fragment of *HMSG11* under the same conditions as above. One clone that hybridized strongly to the probe was subcloned into the Bam H1 site of pBluescript II (Stratagene, La Jolla, CA). This 12,792 bp clone (GenBank Accession No: AF038556) contained three full-length and one partial *MSG* sequences in a head to tail tandem arrangement, similar to what has previously been reported (Garbe and Stringer (1994) *Infect. Immun.* **62**:3092-3101; Stringer *et al.* (1993) *J. Eukaryot. Microbiol.* **40**:821-826). One of the full-length *MSG* sequences did not have a complete open reading frame due to a frame shift between bases 6290 and 6347. The codon corresponding to a methionine at the beginning of rat *P. carinii MSG* clones encoded a valine in all the open reading frames, consistent with earlier observations (Garbe and Stringer (1994) *Infect. Immun.* **62**:3092-3101; Stringer *et al.* (1993) *J. Eukaryot. Microbiol.* **40**:821-826). Nucleotide sequencing was performed as above.

Example 2: Characterization of Human-*P. carinii* MSG Proteins

Figure 1 shows an alignment of the predicted proteins encoded by the full length *MSG* genes cloned by PCR (MSG11, 14, 32, 33, and 35) and Southern (MSGp1 and p3), together with previously published a human (Garbe and Stringer (1994) *Infect. Immun.* 62:3092-3101) and rat *P. carinii* MSG sequence (GenBank accession number L05906). Among the human-*P. carinii* MSG sequences, there is substantial variability downstream of the amino-terminus, while the region near the carboxyl terminus is highly conserved. For example, there is 63% identity in the last 100 amino acids among all the genes (excluding the region encoded by the PCR primer JK152), which is about five times as high as the conservation among the first 100 amino acids (13% excluding the primer region corresponding to primer JK151). Like most known genes of *P. carinii*, all human-*P. carinii MSG* genes show a strong AT bias, especially in the third position (approximately 70% A or T) (Edman *et al.* (1989) *Proc. Natl. Acad. Sci.* USA. **86**:8625-8629; Garbe and Stringer (1994) *Infect. Immun.* 62:3092-3101; Kovacs *et al.* (1993) *J. Biol. Chem.* **268**:6034-6040; Wada *et al.* (1993) *J. Infect. Dis.* **168**:979-985). As in other

MSG molecules, cysteine residues of the human-*P. carinii* MSG molecules are relatively numerous (5.7 to 5.9%) and are highly conserved: 96% of all the cysteine residues present in the human-*P. carinii* MSG clones are conserved in all the clones. When comparing HuMSG11 to rat *P. carinii* MSG clone GP3, 94% of cysteine residues are conserved. The cysteine residues are unevenly distributed in four main regions and often show a pattern of two cysteines separated by 6 to 7 amino acids, similar to what is seen in rat *P. carinii* (Kovacs *et al.* (1993) *J. Biol. Chem.* **268**:6034-6040). There is no predictable pattern to the intervening amino acids. All human MSG proteins share a highly conserved amino acid domain rich in threonine and serine residues near the carboxyl terminus. Seven to thirteen potential N-linked glycosylation sites (NXS/T) were observed in the MSGs. A premature stop codon was seen in MSG 32 after residue 1008 which is most probably due to a PCR artifact resulting in a point mutation; studies using the ligase chain reaction with primers specific for the mutation supported this conclusion.

A. Construction and expression of full length recombinant human-*P. carinii* MSG

The full-length *HMSG32* gene, which contains the premature stop codon, was inserted into pBlueBacHis2A (Invitrogen, Carlsbad, CA) at the Eco R1 site for expression in a baculovirus insect cell system. Correct insertion was confirmed by restriction mapping and sequencing. Isolation of recombinant virus, plaque purification and amplification of high titer virus stock were performed according to the manufacturer's protocols (Invitrogen, Carlsbad, CA). PCR amplification using gene-specific primers was used to confirm the presence of the gene in the virus. Sf9 cells were grown at 27°C in SFII-900 medium (GIBCO BRL Grand Island, NY) with 5% fetal calf serum to a density of 2.0×10^6 cells/ml. Cells were infected at a multiplicity of infection (moi) of 5. Seventy-two hours after infection, cells were harvested by centrifugation, washed with phosphate buffered saline supplemented with PMSF (1 mM/ml), then resuspended in 10 mM Tris-HCl, pH 8 with 1 mM PMSF, and sonicated. The cell lysates were analyzed by SDS-PAGE and western blotting.

SDS-PAGE and western blotting were performed using standard techniques (see Kovacs *et al.* (1988) *J. Immunol.* **140**:2023-2031). Electrophoresis was done in pre-poured discontinuous 8% and 14% acrylamide tris-glycine gels (Novex, San Diego, CA). Proteins were stained by Coomassie blue or transferred to nitrocellulose membranes, following which western blots were performed with a variety of antisera using standard techniques (Kovacs *et al.* (1988) *J. Immunol.* **140**:2023-2031). Recombinant rat *P. carinii* HMSGp3 protein (expressed in a baculovirus system) (Mei *et al.* (1996) *J. Eukarot. Microbiol.* **43**:31S) and purified recombinant β-galactosidase (expressed in the pET 28-*E. coli* system) were used as controls in western blotting.

Anti-peptide antisera were commercially generated in rabbits to a peptide specific for HMSG32 (KMYGLFYGSGKEWFKKLLEKIM (SEQ ID NO: 25), corresponding to amino acids 461-482) and to a conserved human-*P. carinii* MSG epitope contained within the recombinant carboxyl terminal fragment (TITSTITSKITLTST (SEQ ID NO:26) corresponding to amino acids 968 to 982 of MSG32) by the multiple antigenic peptide system method (Posnett *et al.* (1988) *J. Biol. Chem.* **263**:1719-1725) (Research Genetics, Huntsville, AL). Anti-Xpress monoclonal antibody, which detects an epitope tag at the amino terminus of the fusion proteins expressed in pBlueBacHis2A, was purchased from Invitrogen (Carlsbad, CA). T7-tag monoclonal antibody, which detects an epitope tag at the amino terminus of the fusion proteins derived from PET 28A, was purchased from Novagen, Inc. (Madison, WI).

A time course showed that maximal expression occurred after 60-72 hours of infection. The identity of the recombinant protein was confirmed by western blotting using both an antibody against a peptide tag present in the vector as well as an anti-peptide antibody raised against a peptide (SEQ ID NO: 25) specific for MSG32. No reactivity was seen when SF9 cells alone or recombinant baculovirus-derived rat MSG GP3 were used as the targets. Multiple bands were seen in the western blots, especially when using the MSG-specific anti-peptide antibody. These likely represent protein degradation products, or possibly modification of the recombinant protein.

Although rat MSGGP3 could be produced at a high level in a baculovirus system, and was easily purified by affinity chromatograph using a nickel column (Mei *et al.*

(1996) *J. Eukarot. Microbiol.* **43**:31S), prolonged attempts to produce and purify high levels of human-*P. carinii* MSG were unsuccessful.

B. Construction and Expression of the Conserved C-terminal Portion of Human-*P. carinii* MSGs

PCR was used to amplify the conserved carboxy-terminal region of the human-*P. carinii MSG* gene without the carboxyl terminus hydrophobic tail, since this hydrophobic tail could potentially interfere with expression and purification. Primers were designed based on the alignment of five new *MSG* genes as well as the published sequence. The sense primer was JK451 (5'-GAA TTC GAT CTG AAG CCT CTG GAG-3') (SEQ ID NO: 23), and the antisense primer was JK452 (5'-TTC TAG AAA CCC ACT CAT CTT CAA-3') (SEQ ID NO: 24). An Eco RI site was added to the sense primer and an Xba I site, which encoded an in frame stop codon, was added to the antisense primer to facilitate subcloning. One μg of plasmid DNA was used for PCR amplification under the same conditions used above for isolation of PCR clones.

The 306 bp PCR product of carboxy-terminal region amplified from *MSG33* was ligated in frame into pET28A (Novagen, Inc. Madison, WI) at the Eco RI site. pET28A is an expression vector in which a histidine tag precedes the insertion site. The presence of a six histidine (hexa-his) sequence in the expressed portion of the vector preceding the insert allows rapid, one-step purification of the recombinant protein by binding to nickel metal affinity chromatography matrix. Restriction mapping and sequencing were performed to confirm correct insertion. Expression was induced in *E. coli* strain BL21 (DE3) using 1 mM IPTG. Recombinant protein was solubilized with 6M urea and purified by affinity chromatography using a nickel column according to the manufacturer's instructions (Novagen, Inc., Madison, WI). The sample was eluted with elution buffer without urea, dialyzed using 0.5X PBS to eliminate imidazole, and lyophilized for storage.

Recombinant protein was analyzed by SDS-PAGE and western blotting as above. High level expression was observed within two hours; no equivalent band was seen using pET 28A without insert under the same conditions. Although the yield was variable from experiment to experiment, typically about 7 milligrams of purified protein was

obtained from a one liter culture of *E. coli*. The identity of the protein was confirmed by immunoblotting using both T7-tag monoclonal antibody and a polyclonal anti-epitope antibody generated in rabbits against an epitope (SEQ ID NO: 26) contained within the recombinant carboxyl terminal fragment. No reactivity was seen with preimmune rabbit serum, with uninduced *E. coli* extracts, or with second antibody alone.

C. Evaluation of Human Sera Using Antibodies to Human-*P. carinii* MSG

Human sera evaluated by immunoblotting included sera from both AIDS and non-AIDS patients with and without a history of *P. carinii* pneumonia, as well as healthy individuals. Samples included those from 11 immunosuppressed patients with recent or acute *P. carinii* pneumonia but without HIV infection, 5 patients with HIV infection and *P. carinii* pneumonia, 17 patients with HIV infection but without *P. carinii* pneumonia, 3 patients with neither HIV infection nor *P. carinii* pneumonia, and 13 healthy laboratory workers. Human sera were tested at a dilution of 1:100. Horseradish peroxidase-conjugated goat anti-human IgG, alkaline phosphatase conjugated goat anti-rabbit IgG and goat anti-mouse IgG (all from GIBCO BRL) or horseradish peroxidase conjugated goat anti-cat, anti-rat, and anti-mouse IgG (Jackson ImmunoResearch Laboratories, Inc., West Grove, PA) were used as second antibodies in western blotting.

All 49 samples reacted by immunoblotting with the recombinant peptide. Because the recombinant peptide included a vector-derived region, a subset of 4 samples was simultaneous evaluated for reactivity with recombinant β-galactosidase expressed in the same vector. None of the samples reacted with the recombinant β-galactosidase, demonstrating that the reactivity seen was against the *P. carinii* derived peptide region. In addition, little or no reactivity was seen when using rat, mouse, or cat serum.

Example 3: Detection of Human- *P. carinii* Nucleic Acid Sequences.

A. Preparation of a Vector Comprising A Control Sequence

A mimic amplification construct containing a positive control sequence was prepared using the tetracycline resistance (tet^R) gene coding sequence from pBR322 (Backman and Boyer (1983) *Gene* **26**:197). In order to generate a tet^R gene-based amplicon that could be amplified using *MSG*-specific primers JKK14/15 and JKK17, bipartite primers were generated with two distinct annealing regions. The 5' region of each primer was taken from the *MSG* target sequences (*e.g.*, SEQ ID NOS: 17 and 20). The 3' region of each primer was designed to be specific to the tet^R coding sequence. Amplification using these primers generated an amplicon containing an approximately 280 base internal fragment of tet^R coding sequence, with 25 nucleotide *MSG*-specific ends. For amplification, 1 μg of tet^R coding sequence DNA was added to a 50 μl reaction containing primers (25 pM each), dNTPs (0.2 mM), 5 U of AmpliTaq (Perkin-Elmer), and $MgCl_2$ (2.5 mM). The DNA amplification was performed on a Perkin Elmer Cetus DNA thermal cycler. An initial denaturation cycle (2 minutes at 94°C) was followed by 34 cycles of denaturation at 94°C for 1 minute, annealing at 68°C for 1 minute and extension at 72°C for 2 minutes, followed by a final extension after the last cycle at 72°C for 5 minutes.

The resultant 294 base pair amplicon was ligated in to the pCR 2.1 vector and transformed into *E. coli* following the manufacturer's procedures (TA cloning Kit, Invitrogen, Carlsbad, CA). Confirmation of the insert was performed through standard cloning and PCR techniques.

B. Collection and Preparation of Clinical Samples

Clinical samples for use in *MSG*-PCR detection of *P. carinii* can be collected in any conventional way. Sputum was collected as described in Bigby *et al.* (*Am. Rev. Respir. Dis.* **133**:515-518, 1986), and Kovacs *et al.* (*NEJM* **318**:589-593, 1988). Bronchoalveolar lavage (BAL) was performed as described in Ognibene *et al.* (*Am. Rev.*

Respir. Dis. **129**:929-932,1984). Oral washes were carried out by having the subject gargle with 50 cc of normal saline for 10-30 seconds and then expectorate the wash into a sample cup (Helweg-Larsen *et al.* (1998) *J. Clin. Microbiol.* **36**:2068-2072). Serum samples were obtained from blood in a conventional fashion. A 200 μL aliquot of serum was used for DNA extraction.

Oral washes, sputum and bronchoalveolar lavages were spun down 3500 rpm for 10 minutes and the supernatant decanted, leaving approximately 1 ml of liquid in which to resuspend the pellet. Samples were transferred to 2 ml microfuge tubes and centrifuge at 10,000 rpm for 10 minutes to remove remaining liquid. A 250 μL aliquot of InstaGene Matrix (BioRad. Cat. #732-6030, Hercules, CA) was added to the pellet and vortexed briefly. The samples were then incubated at 56° C for 20 minutes, vortexed for 10 seconds and incubated at 100° C for 8 minutes. The samples are vortexed again for 10 seconds and centrifuged at 12,000 rpm for 3 minutes; 5 μL of the resultant supernatant was used in each standard 50 μL PCR reaction.

In certain experiments, DNA was extracted from samples prepared as above using the NucliSens Isolation System (Organon Teknika Corp., Netherlands), using the manufacturer's instructions.

C. Conditions for PCR reactions

To minimize contamination, DNA extraction, amplification and product detection procedures were carried out in separate areas of the laboratory, aerosol-barrier pipette tips were used for all reagent transfers, and multiple negative controls were included in each experiment. In order to minimize carry-over contamination from amplified samples, all specimens were irradiated with UV light after completion of amplification to cross-link the IP-10, which reacts with the PCR product to make it unamplifiable while not interfering with detection (Isaacs *et al.* (1991) *Nucleic Acids Res.* **19**:109-116; Rys and Persing (1993) *J. Clin. Microbiol.* **31**:2356-2360).

***MSG* sequence:** For PCR amplification of human-*P. carinii MSG* in clinical samples, the upstream primer used was an equimolar mixture of JKK14 (SEQ ID NO: 17) (corresponding to the residues of 2809-2833 of *HMSG33,* which is also 2845-2869 of *hMSG11*) and JKK15 (SEQ ID NO: 18) (corresponding to the residues of 2836-2860

of *HMSG32*). The downstream primer used was JKK17 (SEQ ID NO: 20) (complementary to the conserved residues 3028-3052 of *HMSG33,* which is also 3064-3088 of *MSG11*). In experiments wherein the amplified product was detected using the DELFIA™ system, the downstream primer was biotinylated at the 5' end to allow specific capture of amplified sequences through the use of streptavidin.

PCR amplification was carried out in standard PCR reaction mixture (50 mM KCl, 10 mM Tris, pH 8.0, 0.01% gelatin, 3 mM $MgCl_2$, 400 μM dNTPs (Boehringer Mannheim), 1 μM each oligonucleotide primer, and 0.025 units/μl of Amplitaq (Perkin Elmer Cetus)). The HRI AmpStop™ system was used to control carry-over contaminations; IP-10 (a psoralen derivative) (4 μg/μl) was added to each reaction to enable UV cross-linking at the end of the amplification cycle, thereby reducing the possibility of cross contaminating of other samples by amplified products (HRI Research, Inc., Concord, CA).

Samples were amplified using one of the following two PCR cycles: (1) an initial denaturation cycle (5 minutes at 94° C) was followed by 44 cycles of denaturation at 94° C for 30 seconds, annealing at 65° C for 1 minute and extension at 72° C for 2 minutes, followed by a final extension after the last cycle at 72° C for 5 minutes; (2) an initial denaturation at 96° C for 1 minute was followed by 43 cycles of denaturation at 95° C for 1 minute, annealing at 65° C for 1 minute, and extension at 72° C for 1 minute, with a final extension time of 10 minutes at 72° C. All specimens were irradiated with UV light after completion of cycling to cross-link the incorporated IP-10.

Mitochondria large subunit rRNA (MRSU): Previously published PCR primers pAZ102-E and pAZ102-H were used to amplify *P. carinii* mitochondrial large subunit rRNA (MRSU) in clinical samples (Wakefield *et al.* (1990) *Mol. and Biochem. Parasitol.* 43:69-76). Primer pAZ102H was biotinylated at the 5' end to allow streptavidin-mediated capture of the amplified product in experiments wherein the amplified product was detected using the DELFIA™ system. The PCR reaction mixture employed was as above. Samples were amplified using one of the following two PCR cycles: (1) an initial denaturation cycle (2 minutes at 94° C) was followed by 40 cycles of denaturation at 94° C for 1.5 minutes, annealing at 55° C for 1.5 minutes and extension at 72° C for 2 minutes, followed by a final extension after the last cycle at 72° C for 5

minutes; (2) an initial denaturation at 96° C for 1 minute was followed by 43 cycles of denaturation at 95° C for 1 minute, annealing at 65° C for 1 minute, and extension at 72° C for 1 minute, with a final extension time of 10 minutes at 72° C.

D. Detection of Amplified PCR Products

Southern Blotting: Standard southern blotting techniques were used to confirm the PCR results (Tables 2 and 3). Following agarose gel electrophoresis, PCR products were transferred to Hybond N+ membranes (Amersham, Live Science, Arlington Heights, IL). Amplification of human-*P. carinii MSG* was detected using probe JKK16 (SEQ ID NO: 19), which corresponds to residues of 2926-2950 of *HMSG33*. Amplification of *P. carinii* MRSU was detected using pAZ102-L2 (Wakefield *et al.* (1990) *Mol. and Biochem. Parasitol.* 43:69-76). Oligonucleotides were labeled with [γ-^{32}P]-ATP by T4 polynucleotide kinase (Ready-to-Go™ Molecular Biology Reagents, Pharmacia Biotech, Denmark). Prehybridization and hybridization were performed overnight at 52° C in 6 X SSPE, 1% sodium dodecyl sulfate (SDS), 10 X Denhardts' solution (Research Genetics, Huntsville, Alabama). Filters were washed at 52° C in 1 x SSPE, 0.5% SDS for 30 min, then 0.1 x SSPE, 0.5% SDS for 15 minutes.

Time-Resolved Fluorescence: Time-resolved fluorescence detection of amplified sequences was carried out using the DELFIA® system essentially as described by the manufacturer (EG&G Wallac Co.). Using standard procedures, amplicons with incorporated biotin were immobilized in streptavidin-coated microtiter plate wells and washed. Europium-labeled JKK16 was used to probe for the presence of amplified *MSG* sequences; europium-labeled pAz102-L2 was used to probe for the presence of amplified RNA sequences. Results are summarized in Tables 4 and 5, in comparison to DFA staining.

F. Comparison of *P. carinii* Detection Methods

Oral wash samples were collected along with sputum, induced sputum or BAL. All samples were evaluated by direct fluorescent antibody (DFA) staining. DFA staining

was performed using a commercially available kit per the manufacturer's instructions (Genetics Systems, Seattle, WA). Oral wash samples were further tested by PCR, using both primer pairs as detailed above. Summarized results from multiple experiments are shown. Table 2 summarizes the results of a comparison between DFA staining and MSG and MRSU PCR amplification of BAL samples. Table 3 shows the results of a similar comparison using oral wash specimens. Table 4 shows the results of the comparison of samples taken via oral wash; results were determined using the Delfia™ hybridization capture system. Table 5 shows the results of the comparison of samples taken from serum; results were determined using the Delfia™ hybridization capture system.

The DFA-/PCR+ samples (Table 4) likely represent true positive results based on PCR amplification of corresponding sputum samples or concordance between the two PCR methods. One patient with PCP diagnosed by BAL had a negative PCR of oral wash and sputum by both methods, and negative DFA of induced sputum. These data suggest that PCR performed on oral washes can be an accurate, non-invasive means of diagnosing PCP.

Table 2: Results of DFA staining compared to MSG and MRSU gene primer PCR amplification in BAL specimens, as measured by Southern hybridization.

	No. of BAL specimens			
	MSG gene primers		*MRSU* gene primers	
Stain Results	Positive	Negative	Positive	Negative
Positive	7	0	6	1
Negative	0	12	0	12

Table 3: Results of DFA staining compared to MSG and MRSU gene primer PCR amplification in oral wash specimens, as measured by Southern hybridization.

	No. of oral wash specimens			
	MSG gene primers		*MRSU* gene primers	
Stain Results	Positive	Negative	Positive	Negative
Positive	4	4	3	5
Negative	3	70	0	73

Table 4: Results of DFA staining compared to MSG and MRSU gene primer PCR amplification in oral wash specimens, as measured by Delfia™ hybridization capture assay.

	No. of oral wash specimens			
	MSG gene primers		*MRSU* gene primers	
Stain Results	Positive	Negative	Positive	Negative
Positive	11	0	9	2
Negative	4	157	3	158

Table 5: Results of DFA staining compared to MSG and MRSU gene primer PCR amplification in blood serum specimens, as measured by Delfia™ hybridization capture assay.

	No. of serum specimens			
	MSG gene primers		*MRSU* gene primers	
Stain Results	Positive	Negative	Positive	Negative
Positive	3	0	2	1
Negative	0	7	0	7

G. Sensitivity of PCR Using Human-*P. carinii MSG*

The sensitivity of the PCR assay was tested quantitatively by serial dilution of DNA isolated from an autopsy lung sample of an HIV-infected patient with *P. carinii* pneumonia (as above). From this DNA preparation, amplified PCR product could be generated with the *MSG* gene primers (JKK14, JKK15 and JKK17) using about as little as 16 fg of genomic DNA containing human-*P. carinii* DNA as the template. This amount indicates that *MSG* gene amplification is about 10 to 100 fold more sensitive than amplification using the large subunit rRNA gene primers (pAZ102-E and pAZ102-H). This calculation is based on total DNA, the vast majority of which is human DNA, not *P. carinii* DNA, since there is no good method for purifying human-*P. carinii* away from the human DNA in a single sample. Amounts of DNA were measured by spectrophotometry.

The foregoing examples are provided by way of illustration only. One of skill in the art will appreciate that numerous variations on the biological molecules and methods described above may be employed to make and use oligonucleotide primers for the amplification of human-*P. carinii* MSG-encoding sequences, and for their use in detection and diagnosis of *P. carinii* in clinical samples. We claim all such subject matter that falls within the scope and spirit of the following claims.